High-resolution NMR studies of structure and dynamics of human ERp27 indicate extensive inter-domain flexibility

Nader T. Amin[*], A. Katrine Wallis[*], Stephen A. Wells†, Michelle L. Rowe[‡],

Richard A. Williamson[‡], Mark J. Howard[‡] and Robert B. Freedman[*]

[*]*School of Life Sciences, University of Warwick, Coventry, CV4 7AL, UK, †Department of Physics, University of Warwick, Coventry CV4 7AL UK, [‡]School of Biosciences, University of Kent, Canterbury, Kent, CT2 7NJ, UK,*

Corresponding author: Robert B. Freedman, School of Life Sciences, University of Warwick, Coventry, CV4 7AL, UK (email r.b.freedman@warwick.ac.uk)

Short title: Structure, dynamics and inter-domain flexibility of ERp27

SYNOPSIS

ERp27 (Endoplasmic Reticulum Protein 27.7kD) is a homologue of protein disulphide isomerase (PDI) localised to the endoplasmic reticulum. ERp27 is predicted to consist of two thioredoxin-fold domains homologous to the non-catalytic b and b′ domains of PDI. The structure in solution of the N-terminal b-like domain of ERp27 was solved using high-resolution NMR data. The structure confirms that it has the thioredoxin fold and that ERp27 is a member of the PDI family. ^{15}N NMR relaxation data were obtained and ModelFree analysis highlighted limited exchange contributions and slow internal motions, and indicated that the domain has an average order parameter S^2 of 0.79. Comparison of the single-domain structure determined here with the equivalent domain within full-length ERp27, determined independently by x-ray diffraction, indicated very close agreement. The domain interface inferred from NMR data in solution was much more extensive than that observed in the x-ray structure, suggesting that the domains flex independently and that crystallization selects one specific inter-domain orientation. This led us to apply a new rapid method to simulate the flexibility of the full-length protein, establishing that the domains show considerable freedom to flex (tilt and twist) about the inter-domain linker, consistent with the NMR data.

INTRODUCTION

The lumen of the endoplasmic reticulum is the sub-cellular compartment responsible for folding and quality control of proteins exported to the cell surface or extra-cellular spaces. This compartment contains many resident proteins which carry out these folding and quality assurance functions. One prominent group of such resident proteins is the protein disulphide isomerase (PDI) family of proteins (1-3). There are 20 known members of the human PDI family and their best-known role is in facilitating formation of the correct pairing of disulphide bonds, a key aspect of the folding process for most extracellular or cell-surface proteins. Proteins of the PDI family have a range of functions in addition to disulphide bond isomerisation and not all members of the family are catalytically active. All proteins of the PDI family have, or are thought to have, at least one domain with significant structural similarity to thioredoxin (trx). These trx-like domains have a mixed β/α fold with the typical secondary structure topology β_1-α_1-β_2-α_2-β_3-α_3-β_4-β_5-α_4 (3). Those proteins which are catalytically active in disulphide bond isomerisation act through a conserved Cys-Xaa-Xaa-Cys active site motif in a trx-like domain. This motif is located in the N-terminus of α_2 and incorporates the loop immediately preceding this helix. There is typically also a proline residue in the *cis* conformation in the loop between α_3 and β_4. This loop is adjacent to the active site in the tertiary structure and the *cis*-proline is thought to play a role in substrate binding (4). Those members of the PDI family which are not catalytically active possess at least one trx-like domain that lacks the active site motif and may or may not possess the conserved *cis*-proline.

PDI is composed of four trx-like domains; **a, b, b'** and **a',** with an acidic α-helical extension on the C-terminus (2, 5). The **a** and **a'** domains contain the catalytic sites involved in thiol-disulfide oxidoreduction and show high sequence similarity to one another and to thioredoxin. The **b** and **b'** domains are non-catalytic; the **b** domain is of unknown function while the **b'** domain plays a major role in substrate binding (6). This domain contains a hydrophobic pocket which is thought to act as the principal substrate binding site on the protein (7, 8), with the other domains making smaller contributions to substrate binding.

ERp27 is a 27.7 kD endoplasmic reticulum (ER)-resident protein of unknown function and a putative member of the PDI family (9). Based on sequence analysis, ERp27 is thought to be composed of two non-catalytic PDI-like domains, **b** and **b'**. The protein contains two cysteine residues, one in each domain. These cysteines have been shown to exist as free thiols and to be buried within each domain (9). The protein is thus unlikely to be involved in thiol-disulphide oxidoreduction. The C-terminal domain of ERp27 shows sequence similarity to the **b'** domain of PDI, and the similarity includes residues involved in the substrate binding site of PDI (7, 8), suggesting that ERp27 may possess a similar binding site. Indeed, this domain of ERp27 has been shown to bind the 14-residue peptide Δ-somatostatin (9). ERp27 has also been shown to interact with ERp57, another member of the PDI family, both *in vitro* and *in vivo* (9). The binding site for ERp57 has been mapped to the C-terminal (**b')** domain of ERp27.

Here we describe the solution structure and backbone dynamics of the N-terminal (**b)** domain of human ERp27. Structural data have been deposited in the Protein Data Bank (ID 2L4C). NMR relaxation data were collected for this domain, indicating regions of increased mobility. A comparison of the HSQC spectrum for this isolated domain and that for full-length ERp27 identified a large contiguous surface on the **b** domain affected by connection to the **b'**

domain. This putative interface region was considerably more extensive than that found in the x-ray structure of full-length ERp27 (H.Schindelin & F.-X.Kober, personal communication) and stimulated an exploration of the flexibility of ERp27 about the inter-domain linker. We found that the domains show great inter-domain flexibility making extensive transient inter-domain contacts, consistent with the NMR data. This implies that crystallization selected a specific inter-domain orientation from an ensemble of possible orientations.

EXPERIMENTAL

Protein expression and purification

Mature human ERp27 (residues E26-L273) and the individual **b** domain (residues E26-L141) were expressed in *E. coli* BL21 (DE3) pLysS with an N-terminal His tag (MHHHHHHM) as described previously (9). ^{15}N- and ^{15}N/^{13}C-labelled protein was expressed in M9 medium containing $^{15}NH_4Cl$ (1 g/l) or $^{15}NH_4Cl$ (1 g/l) and ^{13}C-glucose (2 g/l). The recombinant protein was purified by immobilised metal affinity chromatography followed by ion-exchange chromatography. Purified protein was concentrated and buffer-exchanged into 25 mM NaH_2PO_4, 100 mM NaCl, pH 6.5. NaN_3 (0.01% w/v) was added as a preservative.

Multidimensional NMR spectroscopy

Protein samples were prepared to a concentration of 1-1.5 mM for structural determination experiments and 0.5 mM for relaxation studies in 25 mM NaH_2PO_4, 100 mM NaCl, pH 6.5. The samples were placed in 5 mm Shigemi BMS-005V tubes with 10% D_2O. All NMR experiments were carried out on four-channel Varian UnityINOVA NMR spectrometers with ^{1}H resonance frequencies of 600 MHz, at the University of Kent, or 800 MHz at the Medical Research Council National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA. All spectra were acquired at 25 °C. Referencing for the ^{1}H chemical shift was based on the position of the water ^{1}H resonance, according to its known relationship with temperature (10) and ^{13}C/^{15}N referencing was achieved using gamma ratios. NMR experiments utilised WATERGATE (11) for solvent suppression to attenuate the ^{1}H water signal. Indirect dimensions were acquired using the hypercomplex method (12). Data were processed on a Linux platform (SuSe 10.0) using NMRPipe (13), and analysed using the CcpNmr Analysis package (14).

NMR Resonance assignment

^{1}H-^{15}N HSQC, CBCA(CO)NH and CBCANH standard triple-resonance NMR experiments were used to assign the backbone Cα, Cβ and amide ^{15}N and ^{1}H resonances. ^{15}N-edited TOCSY-HSQC and HCCH-TOCSY experiments, with mixing times of 60 ms and 18 ms respectively, were used to assign the aliphatic side-chain ^{1}H and ^{13}C resonances. Where necessary, 3D ^{15}N- and ^{13}C-edited NOESY spectra were used to aid in the assignments. Side-chain NH_2 resonances of Asn and Gln were assigned using the ^{1}H-^{15}N HSQC spectrum and a ^{15}N-edited NOESY spectrum. Aromatic side-chain ^{1}H and ^{13}C resonances were assigned using 2D ^{1}H-^{1}H TOCSY with 60 ms mixing time, 2D ^{1}H-^{1}H NOESY with 100 ms mixing time, 2D aromatic ^{1}H-^{1}H NOESY with 80 ms mixing time and a 2D selective aromatic ^{1}H-^{13}C HSQC spectrum.

^{15}N NMR relaxation measurements

^{15}N T_1 and T_2 experiments were acquired using developed pulse sequences comparable to those described elsewhere (15). T_1 and T_2 delay times were set as 128, 256 (x2), 384, 512, 640 (x2), 769 and 897 ms and 20, 40, 60 (x2), 80, 100, 120 (x2), 140, and 160 ms respectively. ^{15}N heteronuclear NOE experiments were collected with a relaxation delay of 5 s with and without saturation of the amide protons that was achieved using 120° high power pulses (16). Relaxation times were calculated as the exponential fit of single exponential decays to peak intensity values; $I = I_0\exp(-t/T_x)$ where $T_x = T_1$ or T_2, and I = resonance intensity at time, t. Heteronuclear NOEs were calculated using the expression $\eta = I/I_0$. ModelFree analysis of the relaxation data was carried out using the ModelFree 4.0 suite of programs (17-19) to obtain the optimal parameter fits of S^2 (order parameter of motion), τ_e (internal correlation time) and R_{ex} (exchange broadening). All ModelFree optimization used the most appropriate model with the lowest number of parameters possible to provide an optimal fit to the relaxation data.

Structure Calculations

NOE distance restraints were obtained from 3D ^{15}N-edited NOESY-HSQC and ^{13}C-edited NOESY-HSQC spectra. Both experiments were recorded at 800 MHz with a mixing time of 80 ms. Dihedral angle (ϕ/φ) restraints were generated from backbone chemical shift data using TALOS (20). Hydrogen bond donors were determined from ^{15}N-HSQC experiments involving hydrogen/deuterium (H/D) exchange and hydrogen bond donor-acceptor pairs within secondary structure elements were identified from preliminary structure calculations. Structure calculations initially used only NOE restraints and were refined by subsequent inclusion of the dihedral angle and hydrogen bond restraints. Structures were calculated using ARIA version 2.1 (21) and CNS version 1.2 (22). For assignment of ambiguous NOESY cross-peaks with ARIA, the frequency window sizes were left at the default settings of 0.02 ppm for direct and 0.04 ppm for indirect proton dimensions and 0.5 ppm for heteronuclear dimensions. For each round of structure calculations, 10 iterations were performed. For the first iteration, 50 structures were calculated and the 20 lowest-energy structures were used for calibration and violation analysis. For the next 7 iterations, 20 structures were calculated, with the 7 lowest-energy structures used for calibration and violation analysis. For the 9th iteration, 50 structures were calculated and the 20 lowest-energy structures were used for calibration and violation analysis. For the final (10th) iteration, 100 structures were calculated. For iterations 0 to 9, the violation tolerances were set to 1000.0, 5.0, 3.0, 1.0, 1.0, 1.0, 0.1 0.1, 0.1 and 0.1 Å respectively and the partial assignment ambiguity cut-off values were set to 1.0, 0.9999, 0.999, 0.99, 0.98, 0.96, 0.93, 0.9, 0.85 and 0.8 respectively. For each iteration, the violation threshold was set to 0.5 Å and the maximum number of contributions for partial assignment was 20. Structure calculations with CNS used torsion angle dynamics. Each iteration used 18,000 cooling steps in the simulated annealing protocol fixed in the ratio of 5:4 for the first and second cooling stages, respectively.

*Mapping of the **b/b'** domain interface on the **b** domain of ERp27*

Residues from the **b** domain of ERp27 potentially forming the interface with the **b'** domain were identified using the minimal shift mapping approach (23,24). Minimal shift mapping provides a measure of the chemical shift difference, Δ_{shift}, that corresponds to the minimal, or closest distance, for each **b** peak to any peak in an equivalent full-length ERp27 (**bb')**

spectrum. Therefore, a peak in the spectrum of **b** with a corresponding peak in **bb'** in an identical position would score a Δ_{shift} of zero with the value of Δ_{shift} increasing as the **bb'** peak moves further away from being coincident to the **b** peak. These changes in chemical shift are indicative of a change in chemical environment in the full-length protein compared to the single-domain protein. Hence, a comparison of the ^{15}N HSQC spectra for the **b** domain and the full-length protein provides identification of residues with NH chemical shifts that are different in the **b** construct compared to the **bb'** construct. For each peak in the HSQC spectrum for the **b** domain, the combined ^{1}H and ^{15}N chemical shift difference was calculated for all peaks in the HSQC spectrum for the full-length protein. The following equation was used: $\Delta_{shift} = \sqrt{(\Delta^{1}H_{N})^{2} + (1/6\ \Delta^{15}N)^{2}}$, where $\Delta^{1}H_{N}$ is the chemical shift difference in the ^{1}H dimension and Δ^{15}N is the chemical shift difference in the ^{15}N dimension. The Δ^{15}N shift difference was reduced by a factor of 6 to account for the difference in ^{15}N-HSQC amide chemical shift ranges of approximately 30 ppm and 5 ppm for ^{15}N and ^{1}H, respectively.

Simulation of the flexibility of full-length ERp27

The atomic co-ordinates of full-length ERp27 were kindly communicated by F.X. Kober and H. Schindelin in advance of publication and were used to simulate and represent the relative motion of the domains. Simulation was by a novel rapid method that analyses the protein structure as a network of rigid clusters and flexible linkers using the program FIRST (25), separately calculates the normal modes of motion of the structure by coarse-grained elastic network modeling using ElNemo (26) and then generates trajectories of motion for the network by geometric simulation using the FRODA algorithm (27) in a calculation in which the normal mode eigenvectors are used to bias the motion of the network (28). This method of simulation of flexible motion retains an all-atom representation of the protein and uses a simplified physical model (29) to maintain the covalent bond geometry and the network of hydrogen bonds and hydrophobic tethers found in the input structure, and to enforce steric exclusion, while neglecting long-range interactions. This simplification allows for a systematic exploration of motion along multiple normal mode directions, achieving large amplitudes, using modest computational resources. The retention of all-atom detail and of the essential local covalent and noncovalent interactions, meanwhile, makes the generated structures physically realistic and useful for the interpretation of experimental data (30).
Normal modes of motion were generated for both full-length ERp27 (molecule B) and for the central **bb'** domains of full-length yeast PDI (from PDB 2B5E). For each structure, trajectories of flexible motion were generated using the ten lowest-frequency non-trivial normal modes; these are modes 7-16, as modes 1-6 are combinations of trivial rigid body motions. Biases were applied both parallel and antiparallel to the mode eigenvectors generating two opposed trajectories for each normal mode, indicated as 7-, 7+. The bias step size in each iteration of the simulation was 0.01 Å. The constraint network included all hydrophobic tethers identified by FIRST in the input structures. Hydrogen bonds identified in the input structure are assigned a (negative) energy in FIRST based on their geometry, and the set of bonds retained in the simulations is determined by setting a cutoff value. Simulations were carried out using cutoff values of -1,-2 and -3 kcal/mol, with essentially identical results. Results are presented for the cutoff of -2 kcal/mol. 2000 conformations were generated along each trajectory with every 100th conformation being recorded as a PDB file for subsequent analysis. All simulations were carried out overnight on a dual-core, 2.4 GHz, desktop workstation running Linux.
In this case our principal interest is in the relative orientation of the two domains, and the

variation of this orientation in the course of large-scale flexible motion. To represent the relative orientations of the two domains, a plane was calculated as representative of the β-sheet plane in each domain, by selecting the Cα atoms of four 'central' residues, namely alternating central residues in each of the adjacent anti-parallel strands β2 and β4. These 4 atoms in each domain give a quadrilateral, from which the plane normals and interplane axis are calculated. The 'tilt' angle between adjacent domains (θ) is defined using the scalar product of the vectors corresponding to the respective plane normals. The dihedral 'twist' angle (ω) makes use of the vector between the 'average' position of the 4 selected Cα atoms in each domain. This vector and the plane normal for each domain define a plane; the dihedral angle between these two planes is the 'twist' angle (ω). Tilt and twist values were extracted for the input structures and for the structures generated in the simulations of flexible motion.

RESULTS

*Resonance assignment of the ERp27 **b** domain*

Excluding the 8 residues of the His-tag (MHHHHHHM), chemical shift assignments were made for 100% of the backbone amide ^{15}N and ^{1}H resonances, 100% of the ^{13}Cα and ^{13}Cβ resonances, 96.7% of the ^{1}Hα resonances, 94.5% of the ^{1}Hβ resonances, 88.8% of the other side-chain ^{1}H resonances (^{1}Hγ, ^{1}Hδ, ^{1}Hε, ^{1}Hζ and ^{1}Hη), 70.6 % of the side-chain ^{13}C resonances (^{13}Cγ, ^{13}Cδ, ^{13}Cε, ^{13}Cζ and ^{13}Cη) and 26.1 % of the side-chain ^{15}N resonances (^{15}Nδ, ^{15}Nε, ^{15}Nζ, ^{15}Nη). Six residues had no side-chain assignments other than ^{13}Cβ; E26, E28, S30, S31, E122 and E125, of which four (E26, E28, S30 and S31) are located at the N-terminus and, following structure determination, were found to be in a flexible extension outside of the core trx fold (see below). The assigned backbone amide resonances are shown in the 2D ^{15}N-^{1}H HSQC spectrum in Figure 1. The well-dispersed pattern of peaks indicates that the protein is folded. A total of 130 HSQC peaks were identified. Of these, 18 were identified as asparagine and glutamine side-chain peaks, one was assigned to the indole amide of the single tryptophan residue, one to the His-tag backbone and the remainder were assigned to the backbone amides of the 110 non-proline residues.

NMR relaxation analysis

^{15}N NMR relaxation parameters T_1, T_2 and heteronuclear NOE were obtained for the **b** domain of ERp27 at 14.1 T (600 MHz ^{1}H) and 25°C (Figure S1). The average values for ^{15}N T_1 and T_2 were recorded as 704 ± 46 ms and 122 ± 64 ms, respectively; these data were used to estimate the global correlation time for the domain of 7.13 ± 0.71 ns using χ^2 optimisation of the theoretical dipolar and practical T_1/T_2 ratios. This estimate of global correlation time confirms that the domain is monomeric under the experimental conditions used. NMR relaxation parameters for each individual ^{15}N were interrogated further using ModelFree analysis that provides motional information in terms of order parameters S^2, internal correlation times τ_e and contributions to chemical exchange R_{ex}, for individual NH vectors across the domain backbone (Figure 2(a-c)). An unstructured N-terminal region was identified from combined low S^2 and high τ_e values for residues 26-38. The average S^2 value across the structured region (residues 39-141) was 0.79 ± 0.04 where S^2 values of 1.0 support a completely rigid domain backbone and 0.01 support an entirely random flexible backbone.

This value of S^2 can be compared to those of oxidised and reduced ERp18 (0.81 and 0.90 respectively) (31) and implies a relatively flexible backbone which may provide capacity for diverse interactions with partner proteins. Average internal correlation times across the structured region were 85 ± 18 ps, a typical value for a well-behaved globular domain. Within the structured region, 5 residues were estimated to have τ_e values above 300 ps as seen in Figure 2(c); L43, V72, V95, G103 and L141. These residues are all associated with loops or are the first residues of secondary structure elements. Their unusually high τ_e values were all obtained with low errors and, according to ModelFree analysis, they indicate significant areas of the domain that have abnormally slow internal motions. Additional ModelFree contribution to R_{ex} were only observed for two residues; L111 and D113 within the loop between strands β4 and β5 (Figure 2(c)).

Structure calculations

In the final round of structure calculations, 1939 NOE-derived distance restraints, 117 dihedral angle restraints and 42 hydrogen bond restraints were used. 100 structures were generated, of which 92 had converged satisfactorily. The converged structures had no distance violations greater than 0.5 Å and no dihedral angle restraint violations greater than 5°. Of the 92 converged structures, the representative structure ensemble was chosen as the 50 lowest-energy structures; superimposed backbone tracings of these structures are shown in Figure 3(a). None of the structures in the ensemble had NOE or hydrogen-bond restraint violations greater than 0.3 Å. The NMR restraints and structural statistics are summarised in Table 1. NOE distribution, structural calculations and NMR relaxation data analysis confirmed that the structured region of the **b** domain extends from residue E39 to L141. Figure 3(a) illustrates the diversity of conformations of the N-terminal region in the individual structures and the structural limits of the domain are also confirmed by comparing the RMSD values between residues 26-141 and those between residues 39-141 (Table 1). Description and analysis of the structure below is restricted to this structured region (E39-L141) and is based on the individual structure in the ensemble closest to the mean structure; this is illustrated in Figures 3(b) and 3(c) showing the assigned secondary structure.

The ordered region of the protein (residues 39-141) consists of a 5-stranded mixed β-sheet surrounded by 5 α-helices (figure 3). The secondary structure topology is $\beta_1\alpha_1\beta_2\alpha_2\beta_3\alpha_3\beta_4\beta_5\alpha_4\alpha_5$. Strands 1, 2 and 3 are parallel to each other while strands 2, 4 and 5 are antiparallel. This corresponds to the variant of the canonical thioredoxin fold (trx-fold) (SCOP reference number 52832) commonly found in the PDI family (3) but with an α-helical insert (α4) prior to the C-terminal α-helix of the fold (α5). The β_1 strand is short (2 residues) but this is common in the family (e.g. in the various domains of yeast PDI (5) this strand comprises 2, 3, 0 and 4 residues). This structural analysis is supported by chemical shift data (Figure S2).

The structure ensemble was analysed using PROCHECK-NMR (32, 33). Over residues 39-141 of the ensemble, 74% of the non-glycine dihedral angles were in the most favoured regions and 26% were in the additionally allowed regions. G103 was the only residue with dihedral angles in a generously allowed region of the plot. This residue is located on the loop between α3 and β4. This corresponds to the *cis*-Pro loop found in many domains in the PDI family. Although there is no conserved proline in this loop in ERp27, G103 is in the equivalent position as the residue preceding the *cis*-Pro in the catalytic PDI domains.

Structural homology

Structural homologues of the **b** domain of ERp27 were identified by a search of the PDB database using DALI (34). 591 structure matches with a Z-score greater than 2.0 were found. Matches were detected to each of the five classes of trx-fold proteins; thioredoxin, glutaredoxin, glutathione-S-transferase, DsbA and glutathione peroxidase. The highest-scoring matches were the structures of the **b** domains of ERp57 (35), human PDI (36), calsequestrin (37), ERp72 (38), yeast PDI (5), ERp29 (39), and ERp44 (40). ERp27 **b** also showed significant structural similarity to thioredoxin despite lacking the active site residues and the conserved *cis*-proline. Ten of the twenty highest-scoring matches were to thioredoxins.

The structure of ERp27 **b** is notable in having an α-helical insert between β5 and α4 of the standard PDI fold. This helix is short, consisting of only 4 residues, and is situated adjacent in the tertiary structure to the loop which contains the active-site (in **a**-type PDI domains). A similar non-canonical helix is found in yeast PDI **b** domain (5). However, the structure of yeast PDI **b** differs significantly from that of ERp27 **b** in that it does not contain α3 of the canonical fold. Both ERp57 **b** and ERp72 **b** differ from the structure of ERp27 **b** most significantly in the loop region between β4 and β5, which is extended in these two proteins. Structure super-positions between ERp27 **b** and selected homologous proteins are shown in Figure 4. A structure-based multiple sequence alignment between ERp27 **b** and selected homologues is shown in Figure S3.

Sequence conservation

The sequence of ERp27 has only been found in vertebrate genomes. The sequence alignment in Figure S4 shows a range of ERp27 **b** sequences from mammal, bird, reptile and fish genomes. The sequence of ERp27 **b** is well conserved across species. The majority of conserved residues are located in the hydrophobic core of the protein, predominantly on β-strands 2, 3 and 4 and on the interior faces of the surrounding α-helices. In the alignment in Figure S4 there are 13 positions that are invariant between ERp27 sequences. These correspond to L43, E57, V58, G62, F63, E68, V85, V95, L108, F109, R110, D113 and F135 of human ERp27. Of these residues, 9 (L43, E58, V58, E68, V85, F109, R110, D113 and F135) are surface-exposed. Residues G62 and F63 are buried within the hydrophobic core of the protein and are situated at the C-terminal end of β2. These two residues are highly conserved in all **b** domains of the human PDI family (Figure S3), with the exception of ERp29. Residue E68 is located on the loop between β2 and α2. This is the location of the active site in the catalytically active domains of the PDI family. E57 and V58 are situated at the N-terminus of β2, R110 is situated at the C-terminus of β4 and D113 is situated on the loop between β4 and β5. These residues are surface-exposed on a common face of the protein. The side-chains of R110 and D113 are positioned so as to form a salt bridge. These residues are also present in the **b** domains of human PDI, ERp57 and ERp29, although in human PDI the arginine is replaced by a lysine (Figure S3).

*The **b** to **b'** domain interface on the **b** domain*

^{1}H-^{15}N HSQC spectra have been collected for both full-length ERp27 and the individual **b** domain (Figure S5). Signals in these spectra are highly sensitive to the local chemical environment of the backbone amide groups. A comparison of these spectra may therefore allow identification of those residues in **b** which are in contact with **b'** in full-length ERp27.

Residues in **b** potentially forming the interface between **b** and **b'** in full-length ERp27 were therefore identified using the minimal shift approach (23, 24). The distance between each assigned cross-peak in the spectrum for the **b** domain to the nearest cross-peak in the spectrum for the full-length protein was measured. The minimal shift values per residue are plotted in the histogram in Figure 5(a). Residues 136-141 at the C-terminus of the **b** domain are inevitably close to the subsequent **b'** domain; omitting these from further consideration, the average minimal shift of the remaining residues was 0.0333 +/- 0.0292. Fig 5(a) shows that the residues with large minimal shifts are clustered in three regions of sequence, namely residues 53-65, 80-87 and 110-119; within these regions, 19 of 31 residues have minimal shifts > 1 standard deviation above the mean and most others have shifts above the mean. To interpret the shifts in relation to structure, all residues within these regions were included in a map of the possible inter-domain surface; these are plotted onto the domain structure in Fig 5(b), which shows clearly that the regions identified in this way are located on one face of the protein. Similar regions of the fold make up the **b** to **b'** domain interfaces in PDI, ERp57, ERp44, calsequestrin and ERp72 (5, 35, 37, 38, 40). Calculation of the electrostatic potential shows the postulated interface region to be predominantly hydrophobic with charged residues around the edge (Figure 5(c)). There are patches of negative charge created e.g. by the surface-exposed side-chains of E57, D113 and E137. There is considerable sequence conservation on the surface of the postulated **b** to **b'** interface (Figure 5(d)). In particular, residues E57, V58, V85, R110, D113 and F135 are strictly conserved in ERp27 (Figure S4), while residues A59, V112, N139 and L141 in the interface are conserved in over 80% of ERp27 sequences.

We were kindly provided with the atomic coordinates of an x-ray structure of full-length ERp27 in advance of publication (Kober, F. X., Koelmel, W., Kuper, J., Drechsler, J., Mais, C., Merhanns, H. M., and Schindelin, H. 'Two binding modes govern the interaction between members of the protein disulphide isomerase family and their substrates') allowing a direct comparison of our inferred interface with that determined by x-ray diffraction.

Our NMR minimal shift data predict that the inter-domain interface is homologous to that between the **b** and **b'** domains in other multi-domain members of the PDI family and this is broadly confirmed by the x-ray structure of the full-length protein, although the x-ray structure indicates an unusual inter-domain angle. For a precise comparison, we selected 'domain contact residues' on the **b** domain by choosing the residues showing the greatest 'minimal shifts' in ^{1}H-^{15}N resonances in the NMR data and the closest inter-domain atom contacts in the crystal structure data, excluding the residues at the C-terminus of the **b** domain (residues 136-141) which are inevitably in contact with or close to the **b'** domain. Both approaches highlighted the same 3 regions on the **b** domain – residues I(53)AATEVAVIGFFQ(65), Q(80)KFPGVSF(87) and R(110)LVDNEQLNL(119). Within these regions, eight residues were identified as contact residues by both criteria (I53, A54, A55, E57, G84, L111, V112, N114), eleven were identified by the NMR criterion only (A59, I61, Q65, Q80, V85, S86, F87, R110, Q116, L117 and L119), and four by the x-ray criterion only (T56, V58, D113 and E115). Hence the contact surface highlighted by NMR is more extensive than that indicated in the crystal structure.

The x-ray data represent a static structure whereas shifts in NMR resonances could represent indirect effects, but will certainly also reflect perturbations that arise from solution dynamics. We therefore attempted to explore the potential flexibility of the molecule using a new rapid method (28).

Simulation of the flexibility of full-length ERp27

Examination of the trajectories of flexible motion showed that, for both ERp27 and PDI **bb'**, the largest amplitudes of flexible motion are achieved by the three lowest-frequency non-trivial normal modes (modes 7,8,9), which are combinations of rotation about the inter-domain axis, and tilting around the inter-domain interface. Further discussion of results is focussed on these modes. To analyse these motions, we have defined the planes represented by the core β-sheet in each domain and extracted the 'tilt' and 'twist' angles between these planes (see Methods). Fig 6(a) is a 'tilt/twist' plot of the inter-domain orientations and confirms that the relative orientation of the domains in ERp27 is distinct from that in yeast PDI (arrows); it shows that the two structures differ in twist angle as well as in tilt angle (ERp27, tilt $+34^{o}$ and twist -33^{o}, compared to PDI **bb',** tilt $+51^{o}$ and twist $+53^{o}$).

Trajectories of flexible motion are shown by calculating and plotting tilt and twist angles for every 100^{th} structure in a simulated trajectory. In the case of ERp27, modes 8 and 9 correspond to tilting motions with the inter-domain link acting as a hinge, (as shown by clear change in tilt angle but very limited change in twist angle); PDI **bb'** shows similar behaviour in modes 7,9 although the change in twist angle is slightly greater. For ERp27, mode 7 - the lowest frequency non-trivial normal mode - shows the most striking results; this mode generates a pronounced twist, effectively a counter-rotation of the two domains around the axis joining the two domains. A similar motion in PDI **bb'** is generated by its normal mode 8. What is notable is that this flexibility is much greater for ERp27 (a total amplitude of twist from -120^{o} to $+ 60^{o}$) than for PDI **bb'**, and that the extent of this motion for ERp27 allows it to take up a relative orientation very close to that found initially in PDI **bb'**. This reorientation of ERp27 to a more PDI-like structure is illustrated in Figure 6(b) and 6(c).

Furthermore, in the course of facile motion along modes 7, 8 and 9, many residues in the loops identified by NMR as 'contact loops' move to make new contacts with the **b'** domain (Fig 6(d), 6(e), 6(f)). Using as contact criterion, that any backbone atom of a residue in the **b** domain is at a distance <7Å from any non-hydrogen atom in the **b'** domain, we find that residues I53-V58, G84-V85 and L111-E115 initially make contact with the **b'** domain in at least one of the 5 molecules of ERp27 in the crystal unit cell. Flexible motion along the trajectories of modes 7-9 brings about new contacts, and almost all the residues identified as being 'contact' residues by the NMR criterion, are now seen to make contact with the **b'** domain. Flexible twisting motion along the mode 7- trajectory (towards the PDI **bb'** conformation) generates contacts between the **b'** domain residues A59, F82, P83, S86 and R110, while motion along trajectory 7+ generates new contacts with P83, F87, R110, Q116, L117, N118. Similarly, tilting motion along mode 8- generates new contacts with A59, V60, Q80, K81, F82, P83, S86, F87 and R110 and that along mode 8+ new contacts with R110, Q116, L117, N118. New contacts made by motion along trajectories 9- and 9+ are all drawn from the same set of residues. Of the 11 non-proline residues in this set, 4 (A59, Q80. S86, L117) are the residues showing the greatest resonance shifts in the NMR analysis (Fig. 5(a)) and these are 4 of 5 residues showing shifts of > mean + 2S.D, while a further 3 of the 'new contact' set (F87, R110 and Q116) have shifts > mean +S.D.

DISCUSSION

The N-terminal (**b**) domain of ERp27 was expressed in high yield, was highly soluble and very stable in solution, allowing the collection of high quality 2D and 3D NMR data. By contrast, the C-terminal domain (**b'**) could not be expressed in isolation as a soluble protein and we did not succeed in strategies aimed at deriving it from cleavage of a fusion construct. Full-length ERp27 was readily expressed and gave a well-dispersed HSQC spectrum (9) but did not remain soluble through long-term NMR data collection at temperatures >30^{o}C, preventing the collection of usable 3D data sets for full assignment and structure calculations. It is now well-established that the ligand-binding **b'**-type domains of the PDI family are the most difficult to generate and study in isolation; high-resolution structures of these domains have only been determined in presence either of an adjacent domain or of a C-terminal extension sequence which is capable of occupying the hydrophobic cleft which constitutes the substrate binding site (41). Without one or other of these extensions in place, the isolated **b'** domains have a tendency to oligomerize and aggregate (42) presumably as a result of the exposure of their hydrophobic ligand binding sites.

The solution structure of the **b** domain reported here is highly congruent with the structure of this domain in full-length ERp27, determined by x-ray diffraction (H. Schindelin & F.-X. Kober, personal communication). The two structures superimpose extremely closely with RMSD values of 0.96 Å over 102 Cα atoms and of 1.02 Å over the corresponding 408 backbone atoms (taking molecule D of the crystal structure as representative). The residues showing the greatest RMSD are in loops at either end of β3 (G103, V112) and in the loop containing the 'additional' helix (E122). The structural context of residue G103 has been noted above as corresponding to the residue preceding the conserved *cis*-Pro in catalytic redox-active trx-domains and this residue also shows up in the relaxation data as having an unusually slow internal correlation time.

It was noted (above) that ERp27 **b** has an α-helical insert between β5 and α4 of the conventional PDI domain fold. The equivalent region in the catalytic **a** and **a'** domains of PDI, the loop between β5 and α4, contains a conserved arginine residue that has been implicated in the catalytic mechanism (43) . Short α-helical inserts also exist in analogous positions in yeast PDI **b** (4) and ERp44 **b'** (40). There is some sequence similarity between these inserts, with the C-terminus of each helix ending in the sequence Asp-X, where X is a hydrophobic branched-chain aliphatic residue. The side-chain of the aspartic acid residue is surface-exposed and that of the hydrophobic residue is directed inwards, towards the core of the protein. The **b**' domain of yeast PDI also contains an α-helical insert at this position, but this is part of a larger insert between β5 and α4 of the fold. There is no structural information available for the **b** domains of PDIp or PDILT, but these proteins appear to have insertions in the protein sequence at a similar position to the helical inserts in ERp27 **b** and yeast PDI **b** (Fig S3). It is therefore possible that PDIp and PDILT also possess the structural feature of an additional α–helix at this location.

We found that ERp27 has a large capacity for flexible motion, due to its two-domain structure with relatively few inter-domain constraints. The lowest-frequency non-trivial normal modes of motion for ERp27 represent a twisting motion around the inter-domain axis (mode 7) and tilting motions where the inter-domain interface acts as a hinge (modes 8 and 9). The **b-b'** moiety of PDI has a similar structure with slightly more inter-domain constraints and a similar (though slightly smaller) range of flexible variation. Twist motion of ERp27

along mode 7- carries its inter-domain orientation from the distinctive state identified in the crystal structure to a PDI-like state. The inference that ERp27 is a highly flexible molecule is corroborated by inter-domain contact information from NMR, which also suggests considerable inter-domain motion (including motion towards a more PDI-like orientational state). Simulations of flexible motion along low-frequency modes generate multiple new inter-domain contacts which correlate well with the inter-domain contacts suggested by NMR. A further implication is that both ERp27 and PDI have considerable conformational flexibility in solution and that the crystal structures have captured particular orientations from a much wider ensemble of flexible variation. The crystallisation of full-length ERp27 in a novel inter-domain orientation has revealed the extent of this variation, not previously revealed by other PDI family crystal structures.

The paucity of information on the function of the **b** domain of ERp27, makes it difficult to interpret these results in functional terms. If the function of the **b** domain is purely passive (e.g. to confer solubility on the functional ligand-binding **b'** domain), then the flexibility of the inter-domain linker could arise negatively from the lack of any functional constraints determining relative orientation; this would contrast with PDI and other multi-domain family members, where the **b-b'** domain pair mustpartially constrain the positions and orientations of the adjacent functional **a** and **a'** domains (4, 44). Alternatively, if the **b** domain plays a role in interactions between ERp27 and other chaperones or folding factors within the ER lumen, then the inter-domain flexibility may be a positive factor in enabling alternative orientations for the domains in interactions with different partners. A recent publication confirmed that many members of the PDI family – including PDI, PDIr, ERp57, ERp72, P5, ERdj5 and ERp29 -- make multiple interactions with partner proteins within the ER ((44) see e.g. their Fig 2A). Unfortunately, ERp27 was not included as a subject in this study, and in any case we lack structural information to underpin the majority of these inferred protein:protein interactions. In the absence of more information on the cellular role of ERp27, further comments on the significance of the flexible inter-domain hinge between **b** and **b'** domains are speculative.

ACKNOWLEDGEMENTS

We are very grateful to Franz-Xaver Kober and Hermann Schindelin (University of Würzburg) for communicating their structure of full-length ERp27 and for very helpful discussions. The majority of NMR data were collected at the University of Kent Biomolecular NMR facility which is supported by the Wellcome Trust. We are also grateful to Drs. Geoff Kelly and Tom Frenkiel (MRC National Institute for Medical Research) for assistance in running 800 MHz NMR datasets.

FUNDING

SAW thanks the Leverhulme Trust for an Early Career Research Fellowship. AKW thanks the Wellcome Trust for a 'Value in People' Fellowship.

REFERENCES

1. Ellgaard, L., and Ruddock, L. W. (2005) The human protein disulphide isomerase family: substrate interactions and functional properties. *EMBO Rep* **6**, 28-32
2. Freedman, R. B., Klappa, P., and Ruddock, L. W. (2002) Protein disulfide isomerases exploit synergy between catalytic and specific binding domains. *EMBO Rep* **3**, 136-140
3. Kozlov, G., Määttänen, P., Thomas, D.Y. and Gehring, K. (2010) A structural overview of the PDI family of proteins. *FEBS J.* **277** 3924-39364.
4. Guddat, L. W., Bardwell, J. C., Zander, T., and Martin, J. L. (1997) The uncharged surface features surrounding the active site of Escherichia coli DsbA are conserved and are implicated in peptide binding. *Protein Sci* **6**, 1148-11565
5. Tian, G., Xiang, S., Noiva, R., Lennarz, W. J., and Schindelin, H. (2006) The crystal structure of yeast protein disulfide isomerase suggests cooperativity between its active sites. *Cell* **124**, 61-736.
6. Klappa, P., Ruddock, L. W., Darby, N. J., and Freedman, R. B. (1998) The b' domain provides the principal peptide-binding site of protein disulfide isomerase but all domains contribute to binding of misfolded proteins. *EMBO J* **17**, 927-935
7. Byrne, L. J., Sidhu, A., Wallis, A. K., Ruddock, L. W., Freedman, R. B., Howard, M. J., and Williamson, R. A. (2009) Mapping of the ligand-binding site on the b' domain of human PDI: interaction with peptide ligands and the x-linker region. *Biochem J* **423**, 209-217
8. Denisov, A. Y., Määttänen, P., Dabrowski, C., Kozlov, G., Thomas, D. Y., and Gehring, K. (2009) Solution structure of the bb' domains of human protein disulfide isomerase. *FEBS J* **276**, 1440-1449
9. Alanen, H. I., Williamson, R. A., Howard, M. J., Hatahet, F. S., Salo, K. E., Kauppila, A., Kellokumpu, S., and Ruddock, L. W. (2006) ERp27, a new non-catalytic endoplasmic reticulum-located human protein disulfide isomerase family member, interacts with ERp57. *J Biol Chem* **281**, 33727-33738
10. Wishart, D. S., and Sykes, B. D. (1994) Chemical shifts as a tool for structure determination. *Methods Enzymol.* **239**, 363-392
11. Piotto, M., Saudek, V., and Sklenář, V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *Journal of Biomolecular NMR* **2**, 661-665
12. States, D. J., Haberkorn, R. A., and Ruben, D. J. (1982) A two-dimensional nuclear Overhauser experiment with pure absorption phase in 4 quadrants. *J. Magn. Reson.* **48**, 286-292
13. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* **6**, 277-293
14. Wim, F. V., Wayne, B., Tim, J. S., Rasmus, H. F., Anne, P., Miguel, L., Eldon, L. U., John, L. M., John, I., and Ernest, D. L. (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins: Structure, Function, and Bioinformatics* **59**, 687-696

15. Barbato, G., Ikura, M., Kay, L. E., Pastor, R. W., and Bax, A. (1992) Backbone dynamics of calmodulin studied by 15N relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. *Biochemistry* **31**, 5269-5278
16. Neuhaus, D., and Vanmierlo, C. P. M. (1992) Measurement of Heteronuclear Noe Enhancements in Biological Macromolecules - a Convenient Pulse Sequence for Use with Aqueous-Solutions. *J. Magn. Resonance* **100**, 221-228
17. Lipari, G., and Szabo, A. (1982) Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity. *J. Am. Chem. Soc.* **104**, 4546-4559
18. Lipari, G., and Szabo, A. (1982) Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .2. Analysis of Experimental Results. *J. Am. Chem. Soc.* **104**, 4559-4570
19. Palmer, A. G., Rance, M., and Wright, P. E. (1991) Intramolecular Motions of a Zinc Finger DNA-Binding Domain from Xfin Characterized by Proton-Detected Natural Abundance C-12 Heteronuclear Nmr-Spectroscopy. *J. Am. Chem. Soc.* **113**, 4371-4380
20. Cornilescu, G., Delaglio, F., and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *Journal of Biomolecular NMR* **13**, 289-302
21. Rieping, W., Habeck, M., Bardiaux, B., Bernard, A., Malliavin, T. E., and Nilges, M. (2007) ARIA2: Automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics* **23**, 381-382
22. Brünger, A., Adams, P., Clore, G., DeLano, W., Gros, P., Grosse-Kunstleve, R., Jiang, J., Kuszewski, J., Nilges, M., Pannu, N., Read, R., Rice, L., Simonson, T., and Warren, G. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr.* **1**, 905-921
23. Waters, L. C., Veverka, V., Bohm, M., Schmedt, T., Choong, P. T., Muskett, F. W., Klempnauer, K. H., and Carr, M. D. (2007) Structure of the C-terminal MA-3 domain of the tumour suppressor protein Pdcd4 and characterization of its interaction with eIF4A. *Oncogene* **26**, 4941-4950
24. Williamson, R. A., Carr, M. D., Frenkiel, T. A., Feeney, J., and Freedman, R. B. (1997) Mapping the Binding Site for Matrix Metalloproteinase on the N-Terminal Domain of the Tissue Inhibitor of Metalloproteinases-2 by NMR Chemical Shift Perturbationâ€ *Biochemistry* **36**, 13882-13889
25. Jacobs, D. J., Rader, A. J., Kuhn, L. A., and Thorpe, M. F. (2001) Protein flexibility predictions using graph theory. *Proteins* **44**, 150-165
26. Suhre, K., and Sanejouand, Y. H. (2004) ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res* **32**, W610-614
27. Wells, S., Menor, S., Hespenheide, B., and Thorpe, M. F. (2005) Constrained geometric simulation of diffusive motion in proteins. *Phys Biol* **2**, S127-136
28. Jimenez-Roldan, J. E., Freedman, R. B., Römer, R. A., and Wells, S. A. (2012) Rapid simulation of protein motion: merging flexibility, rigidity and normal mode analyses. *Phys Biol* **9**, 016008
29. Wells, S. A., and Sartbaeva, A. (2012) Template-Based Geometric Simulation of Flexible Frameworks. *Materials* **5**, 415-431

30. Li, H., Wells, S. A., Jimenez-Roldan, J. E., Römer, R. A., Zhao, Y., Sadler, P. J., and O'Connor, P. B. (2012) Protein flexibility is key to cisplatin crosslinking in calmodulin. *Protein Sci*
31. Rowe, M. L., Ruddock, L. W., Kelly, G., Schmidt, J. M., Williamson, R. A., and Howard, M. J. (2009) Solution structure and dynamics of ERp18, a small endoplasmic reticulum resident oxidoreductase. *Biochemistry* **48**, 4596-4606
32. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* **26**, 283-291
33. Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *Journal of Biomolecular NMR* **8**, 477-486
34. Holm, L., Kaariainen, S., Rosenstrom, P., and Schenkel, A. (2008) Searching protein structure databases with DaliLite v.3. *Bioinformatics* **24**, 2780-2781
35. Kozlov, G., Maattanen, P., Schrag, J. D., Pollock, S., Cygler, M., Nagar, B., Thomas, D. Y., and Gehring, K. (2006) Crystal Structure of the bb' Domains of the Protein Disulfide Isomerase ERp57. *Structure* **14**, 1331-1339
36. Kemmink, J., Dijkstra, K., Mariani, M., Scheek, R. M., Penka, E., Nilges, M., and Darby, N. J. (1999) The structure in solution of the b domain of protein disulfide isomerase*. *Journal of Biomolecular NMR* **13**, 357-368
37. Wang, S., Trumble, W. R., Liao, H., Wesson, C. R., Dunker, A. K., and Kang, C. (1998) Crystal structure of calsequestrin from rabbit skeletal muscle sarcoplasmic reticulum. *Nature Structural Biology* **5**, 476-483
38. Kozlov, G., Määttänen, P., Schrag, J. D., Hura, G. L., Gabrielli, L., Cygler, M., Thomas, D. Y., and Gehring, K. (2009) Structure of the Noncatalytic Domains and Global Fold of the Protein Disulfide Isomerase ERp72. *Structure* **17**, 651-659
39. Barak, N. N., Neumann, P., Sevvana, M., Schutkowski, M., Naumann, K., Malesevic, M., Reichardt, H., Fischer, G., Stubbs, M. T., and Ferrari, D. M. (2009) Crystal Structure and Functional Analysis of the Protein Disulfide Isomerase-Related Protein ERp29. *Journal of Molecular Biology* **385**, 1630-1642
40. Wang, L., Wang, L., Vavassori, S., Li, S., Ke, H., Anelli, T., Degano, M., Ronzoni, R., Sitia, R., Sun, F., and Wang, C.-c. (2008) Crystal structure of human ERp44 shows a dynamic functional modulation by its carboxy-terminal tail *EMBO reports* **9**, 642–647
41. Nguyen, V. D., Wallis, K., Howard, M. J., Haapalainen, A. M., Salo, K. E., Saaranen, M. J., Sidhu, A., Wierenga, R. K., Freedman, R. B., Ruddock, L. W., and Williamson, R. A. (2008) Alternative conformations of the x region of human protein disulphide-isomerase modulate exposure of the substrate binding b' domain. *J Mol Biol* **383**, 1144-1155
42. Wallis, A. K., Sidhu, A., Byrne, L. J., Howard, M. J., Ruddock, L. W., Williamson, R. A., and Freedman, R. B. (2009) The ligand-binding b' domain of human protein disulphide-isomerase mediates homodimerization. *Protein Sci* **18**, 2569-2577
43. Lappi, A. K., Lensink, M. F., Alanen, H. I., Salo, K. E., Lobell, M., Juffer, A. H., and Ruddock, L. W. (2004) A conserved arginine plays a role in the catalytic cycle of the protein disulphide isomerases. *J Mol Biol* **335**, 283-295
44. Jansen, G., Määttänen, P., Denisov, A.Y., Scarffe, L., Schade, B., Balghi, H., Dejgaard, K., Chen, L.Y., Muller, W.J., Gehring, K. & Thomas, D.Y. (2012) An

interaction map of ER chaperones and foldases *Mol.Cell Proteomics* doi:10.1074/mcp.M111.016550

45. Koradi, R., Billeter, M., & Wüthrich, K. (1996) MOLMOL: A program for display and analysis of macromolecular structures *J. Mol. Graphics* **14** 51-55

FIGURE LEGENDS

Table 1. Structural statistics and RMSD values for the ERp27 b domain ensemble (50 structures).

For A, B, D and E, values are reported +/- the standard deviation. A. The energy was obtained using GROMOS96 within DEEPVIEW. B. The RMSD to the mean structure was calculated using MOLMOL (45). The mean structure was calculated independently for residues 26 to 141 and residues 39 to 141. The RMSD over the secondary structure regions used the mean structure calculated for residues 39 to 141. The secondary structure was as assigned using DSSP-CONT. C. The total number of restraints was calculated using the total number of restrained hydrogen bonds, rather than the total number of hydrogen bond restraints. This is because each restrained hydrogen bond had a set of two distance restraints. D. The average number of all restraint violations (NOE, hydrogen bond and dihedral restraint violations) was calculated for all NOE and hydrogen bond violations greater than 0.1 Å and all dihedral angle violations greater than 5°. The data in D and E were generated using CNS. F. Procheck Ramachandran plot statistics calculated for non-glycine residues from 39 to 141 for the ensemble of 50 structures.

Figure 1. Two-dimensional ^{15}N-^{1}H HSQC spectrum of ERp27 b domain.

Peaks are annotated to show the backbone amide assignments. All 110 non-proline residues have been assigned. The indole ε1 amide of the single tryptophan residue (W42) has also been assigned. Side-chain amide NH_2 resonances are connected with horizontal lines.

Figure 2. ModelFree analysis of ^{15}N NMR relaxation data of ERp27 b domain..

Data were collected at 25°C and 14.1 T. Graphical plots illustrate the changes across the sequence of (a) order parameter S^2, (b) internal correlation time τ_e and (c) rate of chemical exchange broadening R_{ex}. A schematic of the domain secondary structure as solved by NMR is shown above each plot.

Figure 3. Solution structure of ERp27 b domain.

(a) Superposition of the ensemble of 50 representative structures, residues 26 to 141. Each structural model in the ensemble was superimposed over the mean structure for the ordered region 39-141. The mean structure was calculated using MOLMOL (45). The Cα traces are shown. The positions of the N and C-termini are labelled “N” and “C”, respectively. (b) and (c) show the single representative structure of ERp27 **b**, residues 39 to 141 and are related by a 180° rotation about the y axis. Secondary structure elements are labelled.

Figure 4. Structure superpositions between ERp27 b and homologous proteins.

Structures were aligned using DALI (34). The single structure of ERp27 **b** with the lowest backbone RMSD to the mean structure over residues 39-141 was used. The structure of ERp27 **b** is shown in blue. The secondary structure elements of ERp27 **b** are labelled. The N- and C-termini of ERp27 **b** are labelled "N" and "C". ERp27 **b** is aligned with (a) human thioredoxin (1ERT, residues 1 to 105, RMSD = 1.8 Å over 95 residues), (b) human PDI **b** domain (2BJX, residues 119 to 216, RMSD = 1.8 Å over 98 residues), (c) yeast PDI **b** domain (2B5E, residues 142 to 237, RMSD = 2.3 Å over 90 residues) and (d) human ERp57 **b** domain (2H8L, residues 135 to 240, RMSD = 1.9 Å over 97 residues). In (a) the catalytic cysteine residues of thioredoxin are displayed as ball-and-stick figures. Residue E68 of ERp27 **b** is also shown. This residue is located in the active site loop and is strictly conserved in ERp27. In (c) the α-helical insertion in ERp27 **b** (α4) and yeast PDI **b** is indicated.

Figure 5. Mapping of the b to b' domain interface on the b domain of ERp27.

(a) Histogram showing the minimal chemical shifts per residue for ERp27 **b** (residues 39-135). Secondary structure is indicated above the histogram. Cylinders represent α-helices and arrows represent β-strands. The minimal shift values corresponding to the mean (μ), mean +1 standard deviation (μ+σ) and mean + 2 standard deviations (μ+2σ) are marked by horizontal lines. (b), (c) and (d) show the contact molecular surface generated using MOLMOL with default settings. In (b), residues are coloured according to the regions of structure showing the highest minimal shift values; residues 53-65 are coloured cyan, residues 80-87are coloured orange and residues110-119 are coloured purple. In (c), the electrostatic potential generated using MOLMOL is shown; blue represents positive charge, red represents negative charge and white represents neutral charge. In (d), sequence conservation is mapped onto the structure; residues with > 80% identity (in the alignment in figure S3) are coloured. (e) Cartoon representation of the protein backbone showing the orientation of the protein used for (b), (c) and (d).

Figure 6. Domain orientation, flexibility and interdomain contacts in ERp27.

(a) Relative orientation of the β-sheet planes in adjacent thioredoxin-fold domains for ERp27 and for the **b-b'** moiety of yeast PDI. The planes are defined by the positions of four Cα atoms in each domain, as follows: for ERp27 domain **b**, V60, G62, I106 and L108, for ERp27 domain **b'**, L164, L166, L222 and I224, for PDI **b** domain, I163, Q165, L202 and I204, for PDI **b'** domain, G260, L262, F314 and I316, Orientation is described by a tilt angle between the plane normals, and by a dihedral twist angle formed by the plane normals and the interplane vector. The orientation in the 2B5E crystal structure of PDI is indicated by a grey arrow and that in the ERp27 crystal structure (9) by a black arrow. Symbols (open for PDI, closed for ERp27) show flexible motion along the three lowest-frequency nontrivial elastic network modes (modes 7,8,9), in positive and negative directions, for each structure. Motion of ERp27 along mode 7- leads to PDI-like interdomain orientations. (b) Backbone cartoon overlay of ERp27 (black) with PDI **b-b'** (white) crystal structures, aligned on the **b'** domain (at left) only, showing very different **b-b'** orientations. (c) Corresponding backbone cartoon overlay of ERp27 after projection along mode 7- (black) with PDI **b-b'** (white) crystal structure; the **b** domain β-sheets of ERp27 and PDI are now coplanar. (d) Backbone cartoon of ERp27 after projection along mode 7- (black). Residues 59 (dark grey), 82,83,86

(light grey), and 110 (white) are shown as spheres. These residues make new contacts with domain b', not found in the input crystal structure. (e) Backbone cartoon of ERp27 after projection along mode 7+ (black), showing residues 110, 116, 117, 118 as white spheres, forming new contacts with domain b'. f) Backbone cartoon of ERp27 after projection along mode 8- (black), showing residues 59,60 (dark grey spheres), 80-83, 86,87 (light grey spheres), 110 (white spheres) forming new contacts with domain b'.

Table l

A. Average energy (kJ mol^{-1})	-2165.152 +/- 121.038
B. Average RMSD to mean structure (Å)	
Residues 26-141	
Backbone	3.27 +/- 0.82
Heavy	3.48 +/- 0.81
Residues 39-141	
Backbone	0.10 +/- 0.02
Heavy	0.54 +/- 0.03
Secondary structure	
Backbone	0.09 +/- 0.02
Heavy	0.52 +/- 0.03
C. Number of restraints	
Total	2098
Average per residue	18.1
NOE	
Total	1939
Ambiguous	1059
Unambiguous	880
Intra (i - j = 0)	296 (33.6%)
Sequential (| i- j| = 1)	232 (26.4%)
Medium-range (1 < |i - j |<5)	146 (16.6%)
Long-range (|i - j| > 5)	206 (23.4%)
Hydrogen bond	
Total constrained	42
Total constraints	84
Long-range constraints (|i - j| > 5)	38
Dihedral angle (φ, ψ)	
Total	117
D. Restraint Violations (mean values per structure)	
All restraints	16.02 +/- 1.73
NOE	
Violations > 0.5Å	0
Violations > 0.3Å	0
Violations > 0.1	13.10 +/- 1.70
Hydrogen Bond	
Violations > 0.5Å	0
Violations > 0.3Å	0
Violations > 0.1Å	2.92 +/- 0.27
Dihedral angle	
Violations > 5°	0
E. RMSD from experimental restraints	
NOEs (Å)	$0.020 +/- 3.505 \times 10^{-4}$
Unambiguous (Å)	$0.018 +/- 6.422 \times 10^{-4}$
Ambiguous (Å)	$0.022 +/- 3.559 \times 10^{-4}$
Hydrogen bonds (Å)	$0.031 +/- 9.481 \times 10^{-4}$
Dihedral angles (°)	$0.110 +/- 1.458 \times 10^{-2}$
F. Procheck statistics (residues 39-141, 50 structures)	
Most favoured regions	3403 (74%)
Additional allowed regions	1197 (26%)
Generously allowed regions	0
Disallowed regions	0

Figure 1

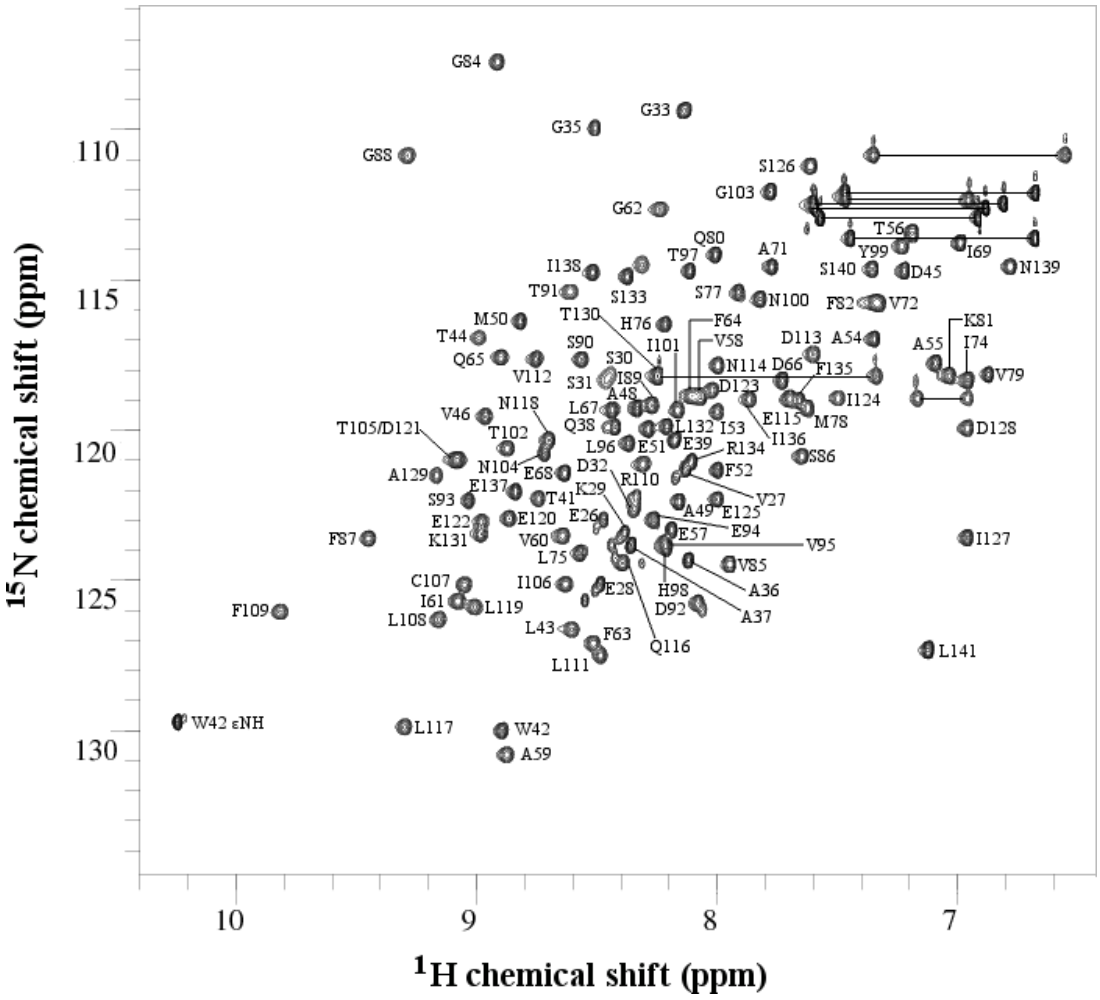

15N chemical shift (ppm)
1H chemical shift (ppm)
110
115
120
125
130
10
9
8
7
G84
G33
G35
G88
S126
G103
G62
T56
Q80
T97
A71
Y99
I69
I138
S140
D45
N139
T91
S133
S77
N100
F82
V72
K81
M50
T130
H76
F64
I74
T44
S90
I101
V58
D113
A54
A55
Q65
S30
N114
D66
F135
V79
V112
S31
I89
D123
I124
A48
N118
L67
V46
Q38
L132
I53
E115
M78
T105/D121
T102
L96
E51
E39
I136
D128
R134
N104
E68
D32
S86
A129
R110
F52
E137
K29
S93
T41
V27
A49
E125
E122
E120
E26
E57
E94
F87
K131
V60
V95
I127
L75
V85
C107
I106
E28
H98
A36
I61
L119
D92
A37
F109
L108
L43
F63
Q116
L111
L141
W42 εNH
L117
W42
A59

Figure 2

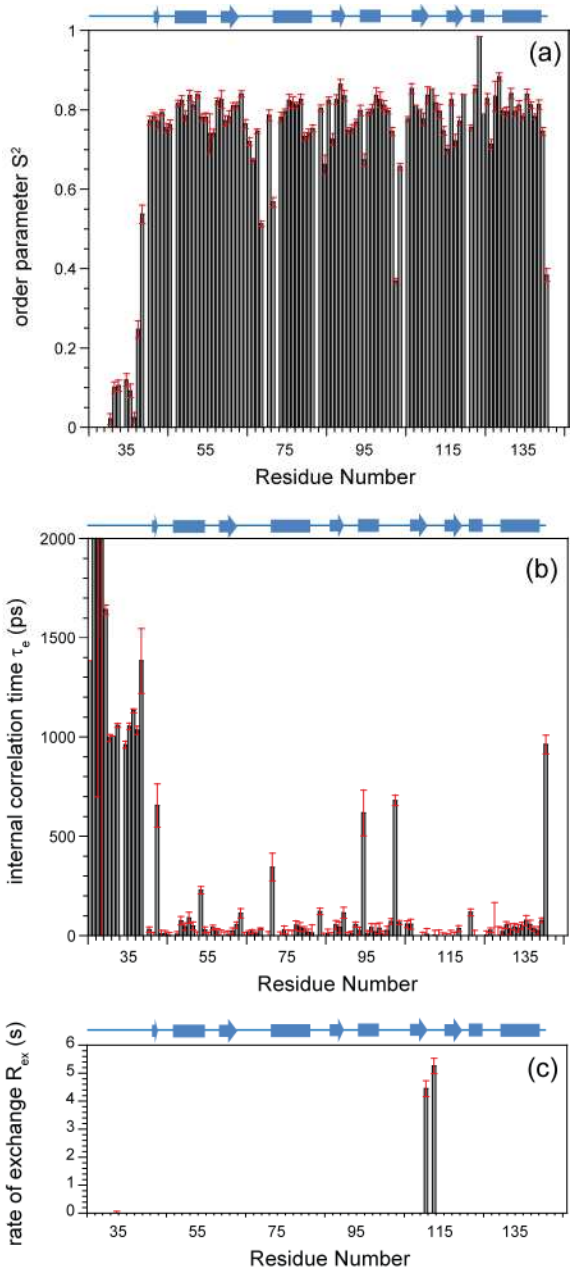

(a)
order parameter S²
Residue Number
(b)
internal correlation time τe (ps)
Residue Number
(c)
rate of exchange Rex (s)
Residue Number

Figure 3

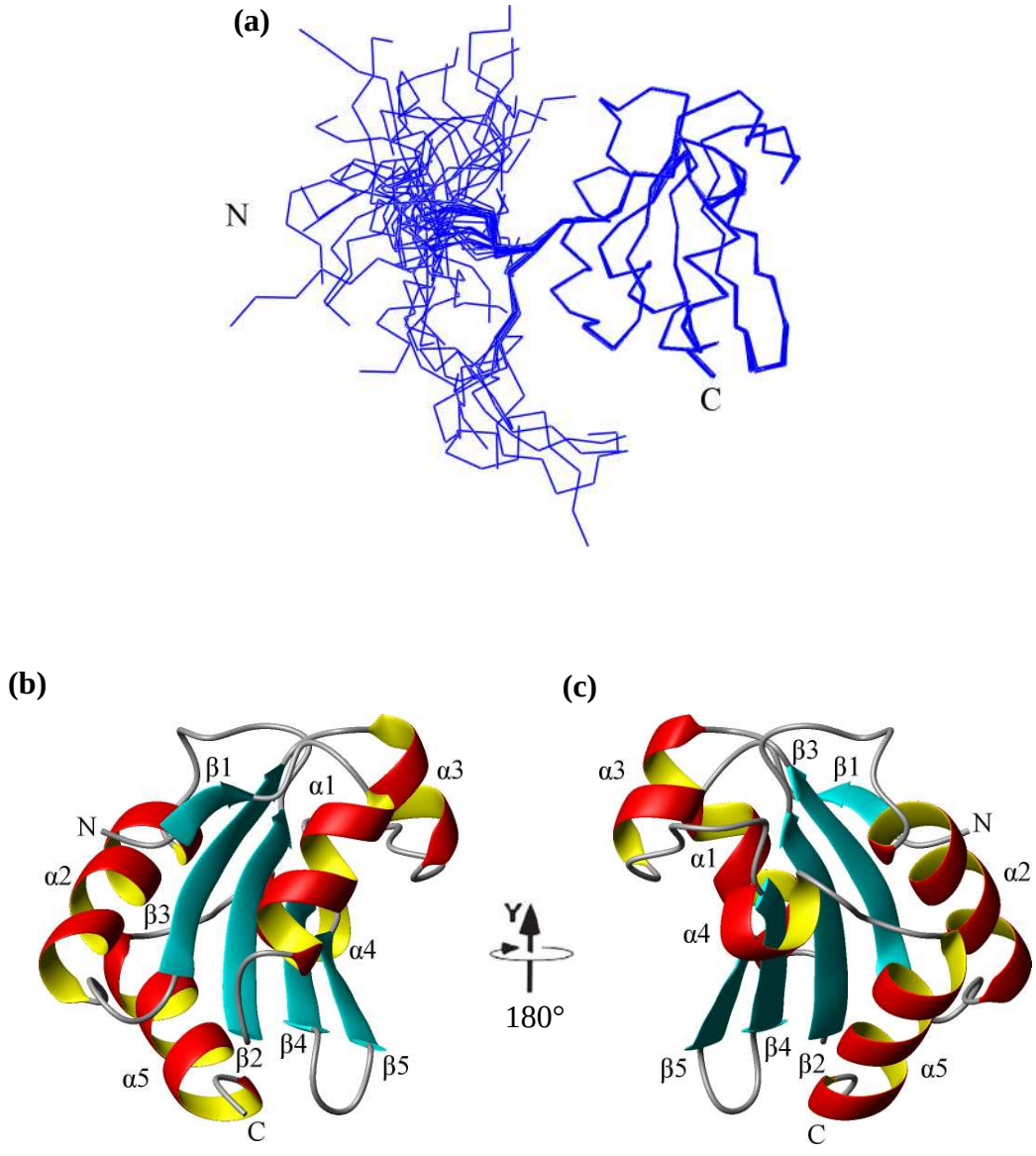

(a)
N
C
(b)
β1
α1
α3
N
α2
β3
α4
β2
β4
β5
α5
C
Y
180°
(c)
α3
β3
β1
N
α1
α2
α4
β5
β4
β2
α5
C

Figure 4

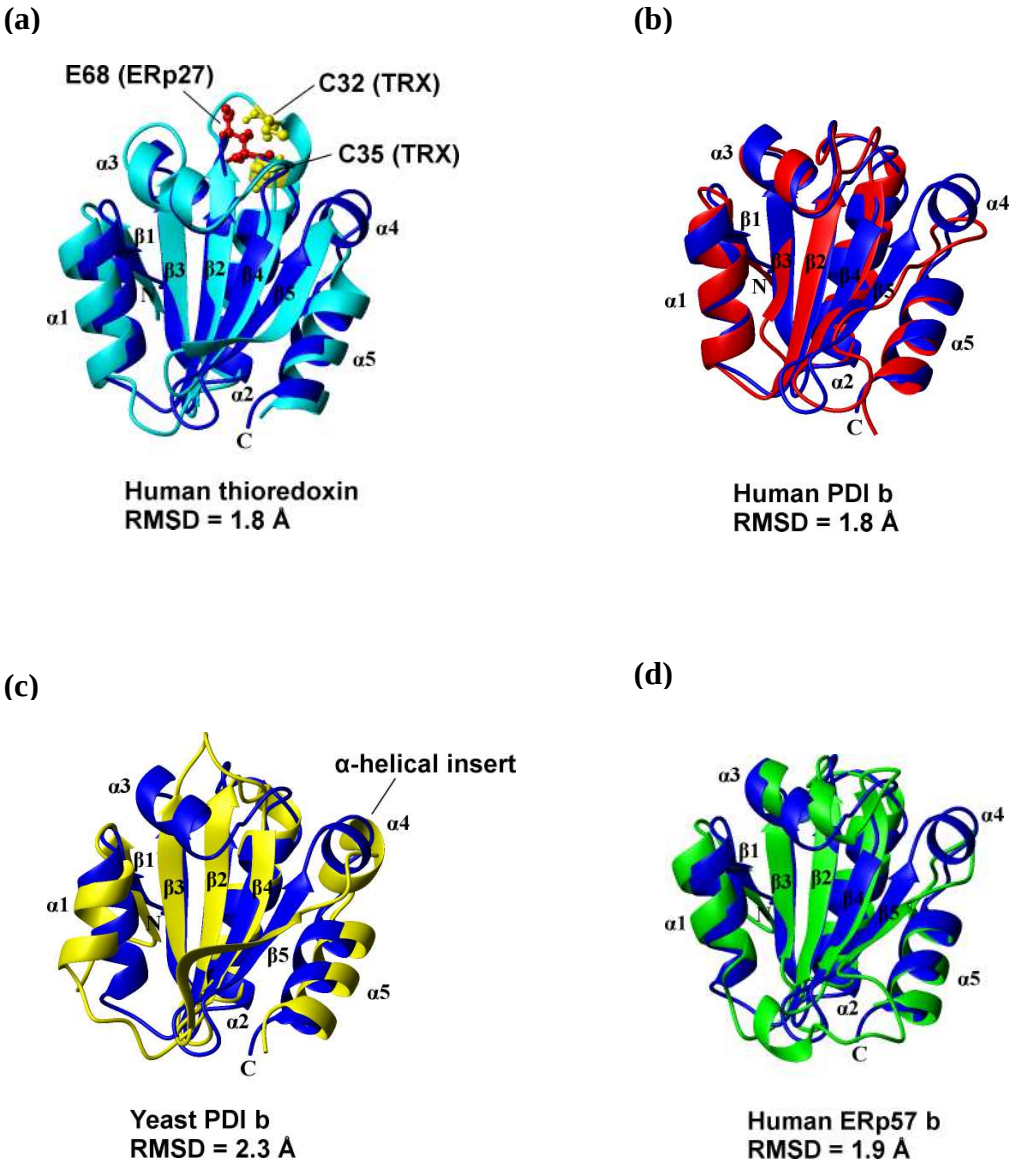
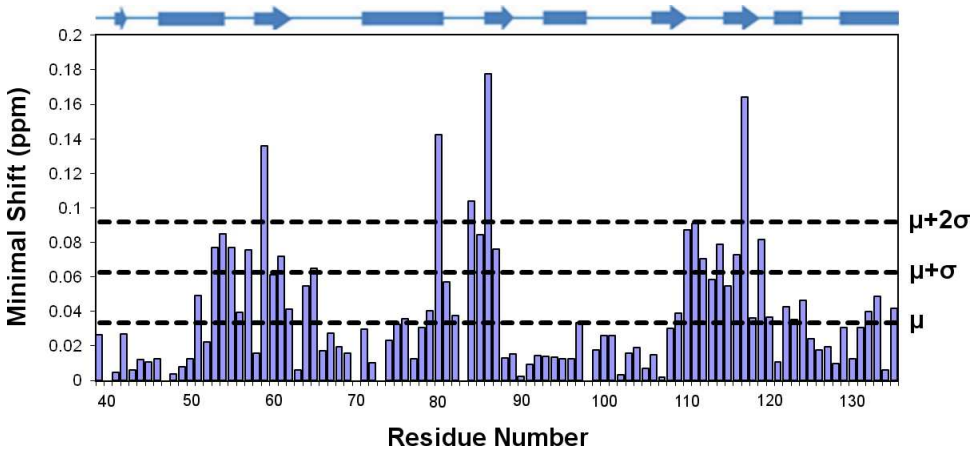
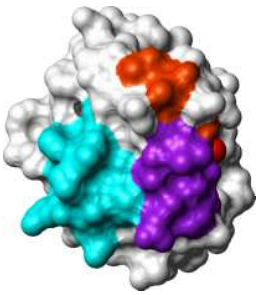

(a)
E68 (ERp27)
C32 (TRX)
C35 (TRX)
α3
β1
α4
β3
β2
β4
β5
α1
α5
α2
C
Human thioredoxin
RMSD = 1.8 Å
(b)
α3
β1
α4
β3
β2
β4
β5
N
α1
α5
α2
C
Human PDI b
RMSD = 1.8 Å
(c)
α-helical insert
α3
α4
β1
β3
β2
β4
α1
β5
α5
α2
C
Yeast PDI b
RMSD = 2.3 Å
(d)
α3
α4
β1
β3
β2
β4
β5
α1
α5
α2
C
Human ERp57 b
RMSD = 1.9 Å

Figure 5

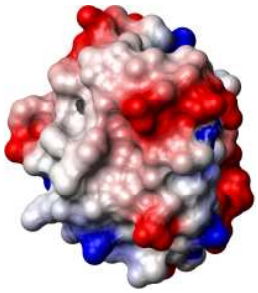
A



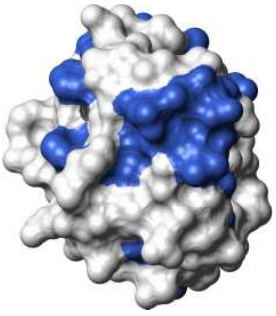
B



C



D



E

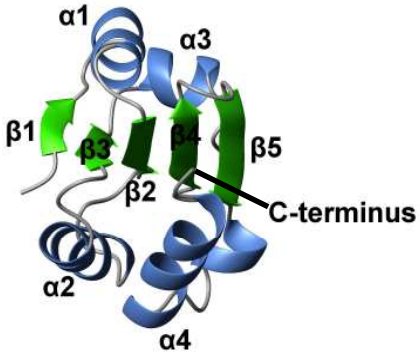
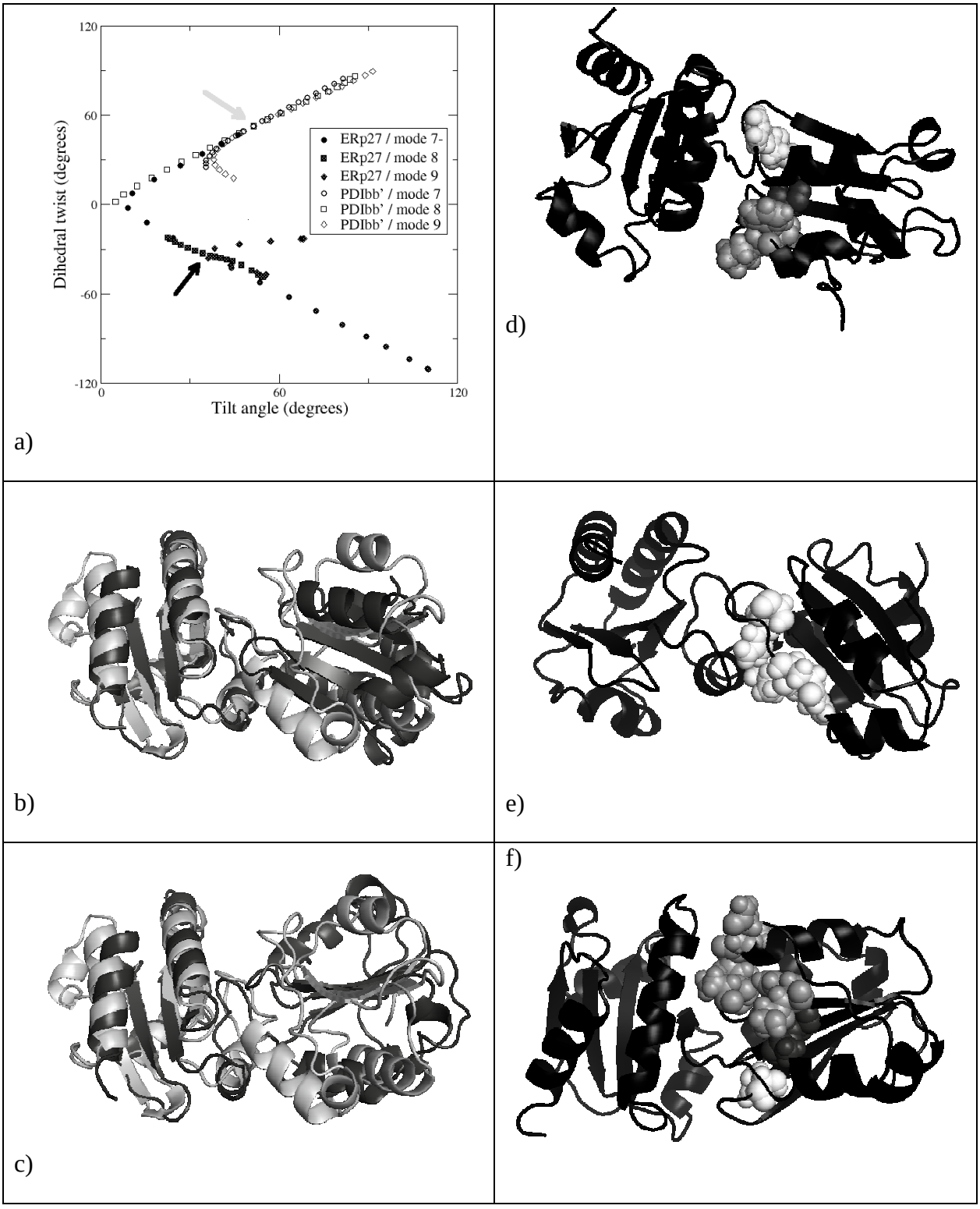


Figure 6


120
60
0
-60
-120
0
60
120
Dihedral twist (degrees)
Tilt angle (degrees)
ERp27 / mode 7-
ERp27 / mode 8
ERp27 / mode 9
PDIbb' / mode 7
PDIbb' / mode 8
PDIbb' / mode 9
a)
d)
b)
e)
c)
f)